J Clin Monit Comput (2014) 28:123–131
DOI 10.1007/s10877-013-9503-0

ORIGINAL RESEARCH

Reliable real-time calculation of heart-rate complexity in critically ill patients using multiple noisy waveform sources

Nehemiah T. Liu · Leopoldo C. Cancio · Jose Salinas · Andriy I. Batchinsky

Received: 4 January 2013 / Accepted: 23 August 2013 / Published online: 30 August 2013

Abstract Heart-rate complexity (HRC) has been proposed as a new vital sign for critical care medicine. The purpose of this research was to develop a reliable method for determining HRC continuously in real time in critically ill patients using multiple waveform channels that also compensates for noisy and unreliable data. Using simultaneously acquired electrocardiogram (Leads I, II, V) and arterial blood pressure waveforms sampled at 360 Hz from 250 patients (over 375 h of patient data), we evaluated a new data fusion framework for computing HRC in real time. The framework employs two algorithms as well as signal quality indices. HRC was calculated (via the method of sample entropy), and equivalence tests were then performed. Bland–Altman plots and box plots of differences between mean HRC values were also obtained. Finally, HRC differences were analyzed by paired t tests. The gold standard for obtaining true means was manual verification of R waves and subsequent entropy calculations. Equivalence tests between mean HRC values derived from manually verified sequences and those derived from automatically detected peaks showed that the "Fusion" values were the least statistically different from the gold standard. Furthermore, the fusion of waveform sources produced better error density distributions than those derived from individual waveforms. The data fusion framework was shown to provide in real-time a reliable continuously streamed HRC value, derived from multiple waveforms in the presence of noise and artifacts. This approach will be validated and tested for assessment of HRC in critically ill patients.

N. T. Liu (✉) · L. C. Cancio · J. Salinas · A. I. Batchinsky
U.S. Army Institute of Surgical Research, 3650 Chambers Pass, Building 3610, Fort Sam Houston, TX 78234-6315, USA
e-mail: nehemiah.liu@us.army.mil

Keywords Signal detection analysis · Electrocardiography · Heart rate · Clinical decision support systems · Machine learning · Automatic data processing

1 Introduction

Heart-rate complexity (HRC) is a method of quantifying the amount of complex variability or irregularity in the heart-rate time series. It is most often obtained by analyzing the R-to-R interval (RRI) of 800 beats or more from a patient's electrocardiogram (ECG). We previously showed that HRC is a sensitive marker of physiologic state during blood loss [1] and trauma [2, 3], and is associated with mortality [2] and the need to perform life-saving interventions in trauma patients [3–5]. Because generation of HRC can be performed remotely and noninvasively and requires small sections of commonly monitored waveforms, HRC can be integrated into a clinical diagnostic tool or decision support system. With today's advances in computing technology, the real-time analysis of HRC in critically ill or injured patients is more realizable than ever.

However, ECG waveforms are often corrupted by artifacts, missing data, and noise that is non-Gaussian and nonstationary. Calculating reliable, real-time HRC values from such signals, and providing confidence intervals for the estimates, is therefore difficult. This is especially true for trauma or critically ill patients. One approach to mitigate this problem is to leverage multiple waveform sources to generate signal quality indices (SQIs) and account for noise and various artifacts.

Leveraging multiple waveform sources to obtain heart-related information has been previously described [6–10], but has not been applied to the development of new vital

Report Documentation Page

Form Approved
OMB No. 0704-0188

1. REPORT DATE **02 APR 2014**	2. REPORT TYPE **N/A**	3. DATES COVERED **-**

4. TITLE AND SUBTITLE **Reliable real-time calculation of heart-rate complexity in critically ill patients using multiple noisy waveform sources**	5a. CONTRACT NUMBER 5b. GRANT NUMBER 5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S) **Liu N. T., Cancio L. C., Salinas J., Batchinsky A. I.,**	5d. PROJECT NUMBER 5e. TASK NUMBER 5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) **United States Army Institute of Surgical Research, JBSA Fort Sam Houston, Tx 78234**	8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	10. SPONSOR/MONITOR’S ACRONYM(S) 11. SPONSOR/MONITOR’S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release, distribution unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT **unclassified**	b. ABSTRACT **unclassified**	c. THIS PAGE **unclassified**	**UU**	**9**	

Standard Form 298 (Rev. 8-98)
Prescribed by ANSI Std Z39-18

signs such as HRC. Owing to their cardiovascular origin, ECG lead waveforms (e.g., limb Leads I and II, or chest Lead V) and pulsatile waveforms [e.g., arterial blood pressure (ABP)] provide independent measures of heart rate that may be suitable for estimation of HRC as well [6, 7]. Furthermore, the ABP is often unaffected by noise, artifacts, and missing data which may degrade the ECG, thereby suggesting that data fusion can provide a more reliable alternative to extracting the heart-rate time series. Nonetheless, the use of multiple waveform channels is only advantageous when the quality of each data source can be determined and the data leveraged accordingly.

The purpose of this research was to develop a reliable method for determining a continuous value of HRC in real time for critical care patients using multiple waveform channels that also compensates for noise and the unreliability of data. The method is based upon the concept of fusing detected peak-to-peak interval (PPI) and RRI estimates derived from multiple noisy waveform sources, such as from ABP and multiple ECG leads, during intensive care unit (ICU) monitoring. This approach automatically rejects degraded waveform data. We hypothesized that a recursive fusion of outputs–first, from several best, published peak detection (PD) algorithms; then, from multiple noisy waveform channels and derived SQIs–could produce a more robust real-time solution for calculating HRC in critically ill patients. By practically fusing the outputs of multiple real-time PD algorithms on multiple waveforms to calculate a streaming HRC value, our data fusion framework may be easily integrated into a real-time HRC software program for decision support and triage in critically ill patients.

2 Materials and methods

The architecture of our framework consisted of multiple real-time PD algorithms applied simultaneously to each waveform source, intermediate logic for fusing detected PPIs and SQIs, and a final block for computing HRC values. The resulting system thus integrated decision stages at the detection and computational levels in order to obtain a final output (see Fig. 1).

2.1 Signals, peak detection, and quality indices

We selected four R-wave detection (RWD) algorithms for real-time implementation based upon (1) their individual performance—as measured using two well-known benchmark parameters [sensitivity (Se), positive predictive value (+P)] against Physionet's Massachusetts Institute of Technology–Beth Israel Hospital (MIT–BIH) Arrhythmia Database and as published in the literature [11–15]—and (2) their ease of implementation for real-time computation. Ease of implementation denoted how well we understood the mechanisms of a detection algorithm. The selected

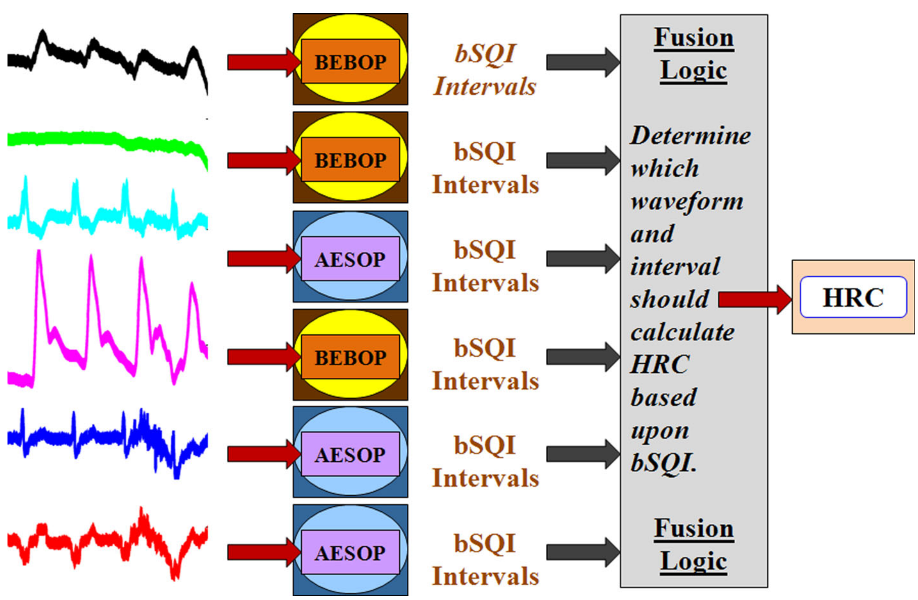

Fig. 1 Architecture of a data fusion framework for calculating heart-rate complexity in real time. Given multiple noisy waveforms at the front-end, the data fusion framework employs instances of two algorithms–one (AESOP) for detecting the R waves of an ECG lead waveform, and the other (BEBOP) for detecting the peaks of a non-ECG waveform—in order to obtain a sequence of peak-to-peak intervals and signal quality indices (bSQI) for each waveform. Upon detection within a specified time window, the framework then employs logic to select a waveform for calculating heart-rate complexity (HRC). A final block performs the actual complexity calculation

algorithms were the Pan–Tompkins [11] (Se: 99.57 %, +P: 99.76 %), Hamilton–Tompkins [12] (Se: 99.69 %, +P: 99.77 %), Christov [13] (Se: 99.74 %, +P: 99.65 %), Afonso–Tompkins–Nguyen–Luo [14] (Se: 99.59 %, +P: 99.56 %), and Zong–Moody–Jiang [15] (Se: 99.65 %, +P: 99.77 %) algorithms; the first two were merged into one component for our final RWD algorithm.

Because RWD algorithms perform differently in different environments [16], we extended the RWD problem to a problem of fusing multiple detection outputs and multiple leads. Tests against animal ECG waveform records suggested that RWD performance could be enhanced by adaptively dropping and re-selecting individual component algorithms and signals based upon performance history and signal quality, respectively. The signal quality index would be an additional output of our final RWD and PD algorithms.

To process multiple waveforms and develop a framework for respective algorithms to operate together in real time, we leveraged signal quality indices for all signals as well as RWD principles to develop new PD algorithms compatible with non-ECG waveforms. Specifically, we modified the final RWD algorithm to produce a final PD algorithm for non-ECG waveforms.

To assess the signal quality of each waveform, we compared the individual performances of multiple PD algorithms on the waveform. Since different detectors are sensitive to different types of noise [16], a comparison of how well algorithms performed within a given time frame provides one estimate of the level of noise in a signal [6, 7]. In this study, concepts from five peak detection algorithms with different noise sensitivities were used. Key concepts of each algorithm are listed in Table 1.

The signal quality of a waveform, with a time frame of N_{Total} beats, was defined in [6, 7] to be the ratio of beats detected synchronously by n PD algorithms to all the beats detected by the final detection algorithm:

$$bSQI = (N_{Matched}/N_{Total}), \tag{1}$$

where $N_{Matched}$ denotes the number of beats (or isolated events) agreed by a specified number n of algorithms, N_{Total} denotes both the time frame and beats detected by the final detection algorithm, and $bSQI$ denotes the waveform's beat SQI. Following fusion at the detection level, our framework selected outputs from the waveform with the highest $bSQI$.

In other words, whenever a chosen number of component algorithms detected the same peak of a waveform, a match was recorded. The higher the number of matches within a specified time window, the higher the beat signal quality index (bSQI). We found this method to be sufficient for comparing the signal qualities of different waveform signals.

Lastly, to make all algorithms platform-independent and operable for real-time output, we implemented them in the Java (Sun Microsystems, Sunnyvale, CA, USA) programming language using the Eclipse Integrated Development Environment (Eclipse Foundation, Ottawa, Canada).

2.2 Heart-rate complexity

We calculated HRC via the method of sample entropy, SampEn(m, r, N), which equals the negative natural logarithm of the conditional probability that two epochs similar for m intervals remain similar at the next interval, given a sequence of N intervals and excluding self-matches. Here, similarity means that peak-to-peak intervals differ by no more than some tolerance r (in milliseconds) [17, 18].

For clarity, sample entropy was computed by the following equations:

$$SampEn(m, r, N) = -\ln(A/B), \tag{2}$$

$$B = [(N-m-1)/2]\sum_{i=1}^{N-m} B_i^r(m), \tag{3}$$

$$A = [(N-m-1)/2]\sum_{i=1}^{N-m} A_i^r(m). \tag{4}$$

In other words, for a sequence of N intervals, if $x_m(i)$ is an epoch of m consecutive intervals starting at index i and running from $i = 1, \ldots, N-m$, then $B_i^r(m)$ denotes the number of epochs $x_m(j)$ within r of $x_m(i)$, for $i \neq j$,

Table 1 Key concepts of selected peak detection algorithms

Algorithm	Key concepts
Pan–Tompkins [11]	Derivative-based signal processing; integer filters; the adaptation of thresholds using recent signal peaks and noise peaks; a search-back mechanism for finding missed beats; refractory blanking; T-wave identification
Hamilton–Tompkins [12]	(See above); fiducial mark placement and consistency; mean peak level estimation; baseline shift discrimination; optimization of search-back detection thresholds
Christov [13]	Combination of three independent adaptive thresholds; search-back mechanism for finding missed beats
Afonso–Tompkins–Nguyen–Luo [14]	Multi-rate signal processing; signal decomposition into sub-bands using a filter bank; feature extraction; single-channel detection blocks and decision levels; the adaptation of detection strengths using signal and noise histories; partial refractory blanking
Zong–Moody–Jiang [15]	Curve length transformation; noise suppression using sign consistency; threshold adaptation; refractory blanking

multiplied by $(N-m-1)^{-1}$, and $A_i^r(m)$ denotes the number of epochs $x_{m+1}(j)$ within r of $x_{m+1}(i)$, for $i \neq j$, multiplied by $(N-m-1)^{-1}$ [18–20].

Parametric values ($N = 200$, $m = 2$, $r = 6$) were established from previous work [1–5]. A higher SampEn implies a more "complex" signal as well as a higher likelihood that the signal belongs to a healthy patient [19–25].

2.3 Patient data and clinical validation

250 ICU patients, as described by records in the Massachusetts General Hospital/Marquette Foundation (MGH/MF) Waveform Database [26–28], were selected for this study. Of these patients, 155 were males, 77 were females, and 18 were not specified. In addition, 20 patients had atrial fibrillation (AF), 65 had sinus tachycardia (ST), 111 had normal sinus rhythm (NSR), and the remaining patients had other rhythms, such as sinus bradycardia or ventricular pacing. Demographics of patients are shown in Table 2.

Selection of the MGH/MF Waveform Database was based upon the following considerations. First, this database was developed to extend the scope of the MIT–BIH Arrhythmia Database [26], which has been historically utilized much for beat detection and in our previous work [29]. We desired the MGH/MF Waveform's Database's similarity to and improvements over its predecessor, such as the availability of simultaneous hemodynamic data and multiple ECG leads. Second, all records were easily accessible and documented, and none were excluded from the study; this would not have been made possible by larger and/or more recent sources. Third, because of the patients' broad demographics (see underlying rhythms in Table 2), we were able to obtain a wide range of HRC values needed for analysis and comparison without filtering records. Lastly, online documentation simplified the task of classifying patients into groups [26]. These considerations made it more suitable for us to choose the MGH/MF Waveform Database over other sources.

Using simultaneously acquired ECG and ABP waveforms from these records, we evaluated our new data fusion framework. Only three ECG leads (Leads I, II, V) were available, and all waveforms were sampled at 375 Hz. The dominant lead was Lead II. Validation involved over 375 h

Table 2 Demographics of 250 patients in the MGH/MF waveform database

Patients	Total		Age		HR		High ABP		Low ABP	
	#	%	Mean	Std	Mean	Std	Mean	Std	Mean	Std
Entire database	250	100.0	58.4	22.0	91.4	19.4	127.3	27.8	58.4	14.7
Gender										
Females	77	30.8	57.1	21.4	91.6	18.6	126.6	28.2	58.8	13.6
Males	155	62.0	59.1	22.4	91.4	19.9	127.6	27.7	58.2	15.2
Unknown	18	7.2	–	–	–	–	–	–	–	–
Underlying rhythm										
Atrial fibrillation	20	8.0	73.3	9.2	91.2	13.6	118.6	24.8	53.8	9.8
Atrial flutter	4	1.6	72.0	6.1	64.3	12.2	147.5	25.0	46.0	4.9
Atrial pacing	5	2.0	66.4	9.7	89.6	8.9	154.8	38.9	56.2	24.8
Atrial tachycardia	1	0.4	0.8	–	–	–	–	–	–	–
Atrioventricular pacing	1	0.4	73.0	–	89.0	–	134.0	–	31.0	–
Chaotic atrial rhythm	2	0.8	70.0	1.4	80.0		126.0	8.5	56.0	8.5
Dual chamber pacing	1	0.4	52.0	–	–	–	–	–	–	–
Junctional escape rhythm	1	0.4	80.0	–	140.0	–	50.0	–	56.0	–
Junctional rhythm	2	0.8	60.0	–	96.0	–	110.0	–	70.0	–
Junctional tachycardia	1	0.4	55.0	–	103.0	–	168.0	–	10.0	–
Multifocal atrial tachycardia	1	0.4	77.0	–	120.0	–	123.0	–	54.0	–
Sinus arrhythmia	1	0.4	76.0	–	66.0	–	160.0	–	80.0	–
Sinus bradycardia	6	2.4	67.3	1.2	50.8	8.2	172.0	27.8	44.0	32.9
Sinus rhythm	111	44.4	13.3	18.8	9.0	10.7	20.8	26.1	9.7	12.4
Sinus tachycardia	65	26.0	50.7	27.2	113.3	11.1	120.9	27.7	59.6	15.4
Unknown	20	8.0	0.2	0.3	–	–	–	–	–	–
Ventricular pacing	8	3.2	62.5	17.8	78.0	21.7	116.3	20.5	60.6	9.4

HR heart rate (beats per minute), *ABP* arterial blood pressure (mm Hg), *std* standard deviation

of patient data describing patients with a broad spectrum of conditions. Individual recordings varied in length from 12 to 86 min, and in most cases, were about an hour long.

Evaluation of the data fusion framework began with calculation of mean entropy values for each set of individual SampEn values corresponding to records of the MGH/MF Waveform Database. We performed data analysis by applying a sliding window of 200 PPIs ($N = 200$, $m = 2$, $r = 6$) to interval sequences either manually verified or detected from one of the waveforms mentioned above (ECG Leads I, II, V and ABP). Thus, we obtained $5 \times 250 = 1{,}250$ mean values, 250 values coming from manually verified sequences in the database. Furthermore, we calculated individual SampEn values across every record using the fusion of ECG Leads I, II, V, and ABP and then obtained mean entropy values. The gold standard for validation was manual verification of R waves, which was accomplished by manually picking times of R waves on time points of the ECG. After hand-picking R waves of all human records, times and RRIs were written to text files for future reference.

2.4 Statistics

We used the parameters Se and +P to compare the detection performances of our data fusion framework against individual waveforms and a combination of the four available waveforms (see Table 3). Further comparisons were made using histogram plots (see Fig. 2). In addition, mean HRC values were calculated for all records, and paired t tests (in which the null hypothesis was that no difference exists between groups) as well as equivalence tests (two one-sided t tests) were then performed in order to compare values derived from automatically detected peaks

Table 3 R-wave detection performance against MGH/MF waveform database

Waveform	Verified	TP	FP	FN	Se (%)	+P (%)	Avg(Se, +P)
Fusion	1,526,672	1,382,804	47,236	143,868	90.6	96.7	93.7
Lead I	1,526,672	1,245,965	297,559	280,707	81.6	80.7	81.2
Lead II	1,526,672	1,433,281	122,752	93,391	93.9	92.1	93.0
Lead V	1,526,672	1,392,478	161,615	134,194	91.2	89.6	90.4
ABP	1,526,672	1,022,140	443,551	504,532	67.0	69.7	68.4

ABP arterial blood pressure, *TP* true positive, *FP* false positive, *FN* false negative, *Se* sensitivity, *+P* positive predictive value, *Avg(Se, +P)* average of Se and +P

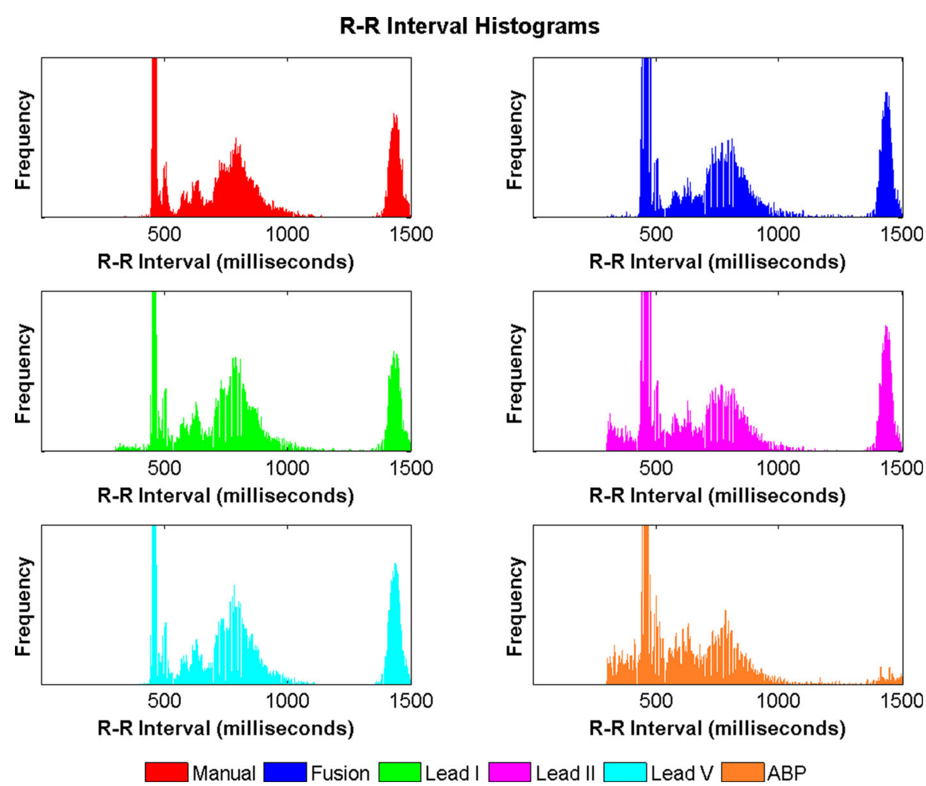

Fig. 2 Histogram plots. Histogram plots for manually verified sequences and detected interval sequences are shown. Detected sequences come from either a fusion of ECG Leads I, II, V, and ABP (Fusion); ECG Lead I; ECG Lead II; ECG Lead V; or ABP

Table 4 *P* values of paired t tests between complexity means (manual vs. detected)

(p value)		Fusion	Lead I	Lead II	Lead V	ABP
	Mean	0.83 ± 0.59	0.88 ± 0.58	0.90 ± 0.58	0.86 ± 0.59	1.43 ± 0.54
Manual	0.84 ± 0.59	$p = 0.06$	$p < 0.001$	$p < 0.001$	$p = 0.02$	$p < 0.0001$

Table 5 *P* values of equivalence tests between complexity means (manual vs. detected)

	Fusion	Lead I	Lead II	Lead V	ABP
Diff in means	0.010	−0.041	−0.055	−0.014	−0.583
Std err of diff	0.053	0.053	0.053	0.054	0.051
Max p value	$p = 0.045$	$p = 0.133$	$p = 0.201$	$p = 0.056$	$p = 1.000$
1-α CI for diff	0.078, 0.097	−0.100, 0.018	−0.143, 0.031	−0.103, 0.074	−0.667, −0.498

Diff difference, *Std err* standard error, *1-α CI* 1-alpha confidence interval, α test size = 0.05, specified difference threshold = 0.1, confidence level = 0.9

("Fusion", Lead I, Lead II, Lead V, and ABP) with those derived from manually verified sequences (see Tables 4, 5, respectively). JMP version 9.0.0 (SAS Institute, Cary, NC, USA) and the R Language (http://www.r-project.org/), a well-known open-source statistical software package, were used for statistical analysis.

For completeness, Bland–Altman plots were also obtained by subtracting "detected" HRC values from the criterion standard (see Bland–Altman plots in Fig. 3; NSR data in green squares, AF data in blue circles, ST data in magenta diamonds, "gold standard" is manually verified peaks). Box plots of differences between mean HRC values were also obtained, i.e., differences corresponding to Bland–Altman plots (see Fig. 3). Finally, differences were analyzed by paired t tests in order to determine whether statistical significances existed between errors from the "fusion" of waveform sources and each of the single waveform sources (see Table 6).

3 Results

We developed the Automated Electrocardiogram Selection of Peaks (AESOP) algorithm to implement the fusion functions for merging ECG peak results from individual algorithms in real time. This fusion algorithm employs four R-wave detectors as inputs and returns final detected peaks, corresponding times, and beat signal quality indices as outputs in approximate real time. Similar to a nearest-neighbor selection scheme, the AESOP algorithm selects the end time corresponding to a mode RRI or the RRI closest to the previous averaged 12 RRIs. In other words, if two or more component algorithms detect the same ECG peak, the AESOP algorithm selects the end time and peak value corresponding to the mode RRI's end time. Otherwise, the algorithm selects the end time and ECG peak yielding an RRI closest to the previous averaged 12 RRIs; this number (12) was chosen based upon heuristics in order to ensure a reasonable average. The AESOP algorithm required less than 6 s to analyze one record of the MIT–BIH Arrhythmia Database on an Intel® Core™ Duo central processing unit at 2.93 GHz [29].

Similarly, we developed the Bypassing Electrocardiogram Beats or Peaks (BEBOP) algorithm for merging non-ECG peak results. This algorithm differs from the AESOP algorithm in two of its sub-implementations, namely, that two tailored versions of the *ATNL* algorithm now replace the *ATNL* and *C* algorithms in the AESOP algorithm, one version for detecting only positive–negative slope deflections in order to focus on systole dynamics and the other for detecting positive–negative slope deflections of a first-order derivative of the original signal.

To avoid biasing this study, we used swine waveform data to systematically tune the values of N_{Total} and n in the *bSQI*. Starting with $N_{Total} = 10$ and incrementing by multiples of 2, and then, by multiples of 5, we determined that $n = 2$, 3 and $N_{Total} = 30$ yielded reasonable indices between 0 and 100 % with 3 % resolution for sampling frequency of 500 Hz. For swine with an average heart rate of 120 beats per min, this corresponded to a time frame of roughly 15 s.

For 250 ICU patient records in the MGH/MF Waveform Database, the data fusion framework (AESOP, BEBOP) achieved an averaged Se and +P of 93.7 %, thereby outperforming results for individual waveforms in terms of mean Se/+P, i.e., tradeoff between Se and +P (see Table 3). In terms of operating points, out of 1,526,672 true beats, the framework detected 1,382,804 TPs, 47,236 FPs, and 143,868 FNs (see Table 3). Histogram plots of PPI sequences are shown in Fig. 2. Importantly, paired t tests (in which the null hypothesis was that no difference exists between mean HRC values derived from manually verified

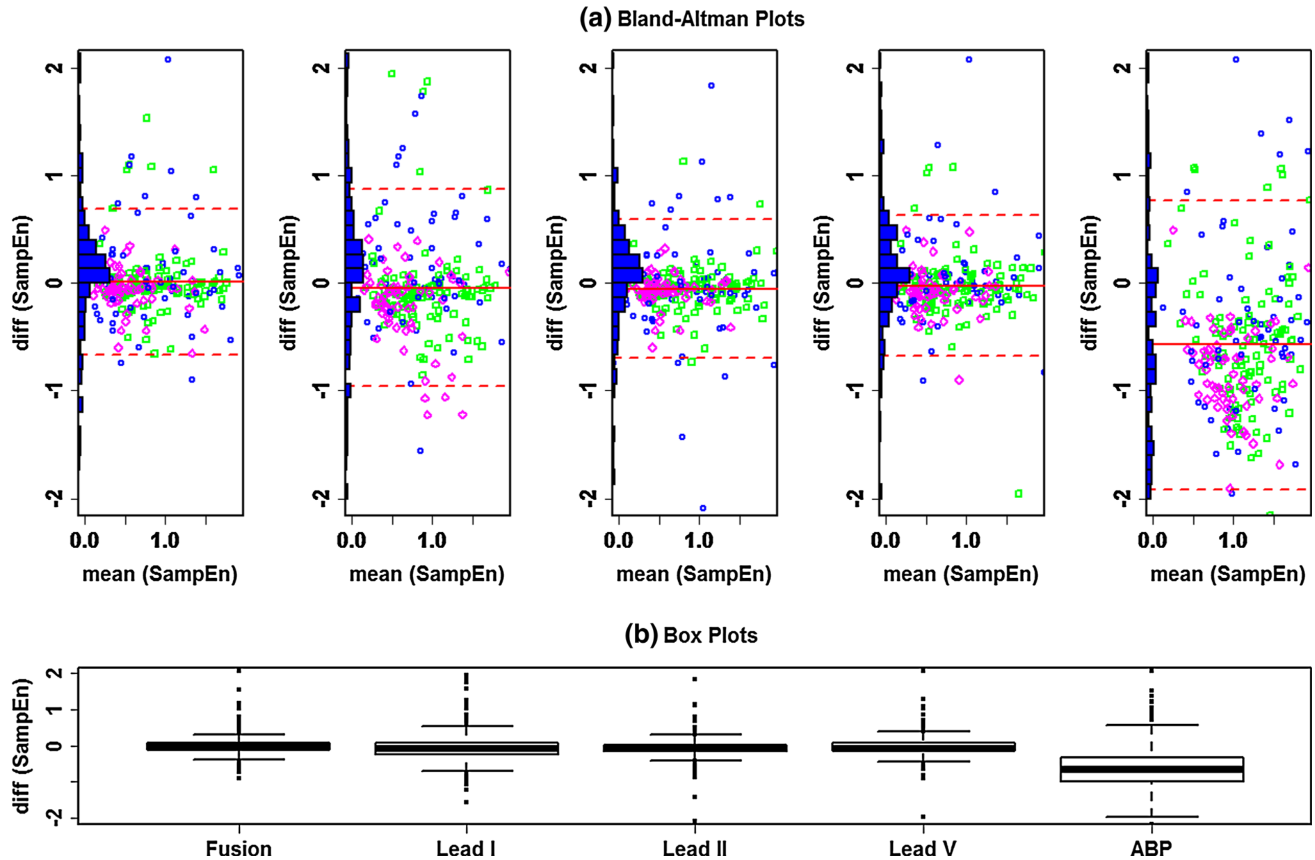

Fig. 3 Bland–Altman and Box-and-whisker plots. **a** Bland–Altman plots for mean entropy values derived from manually verified sequences versus mean entropy values derived from detected interval sequences are shown. Detected sequences come from either a fusion of ECG Leads I, II, V, and ABP (Fusion); ECG Lead I; ECG Lead V; or ABP. Error density distributions (histograms) superimposed on the left-hand sides of each Bland–Altman plot. *Green squares* denote patients with normal sinus rhythm, while *magenta diamonds* denote patients with sinus tachycardia. *Blue circles* denote patients with other underlying ECG rhythms. **b** Box-and-whisker plots. *Box plots* for the differences between mean entropy values derived from manually verified sequences and mean entropy values derived from detected interval sequences are shown. Detected sequences come from either a fusion of ECG Leads I, II, V, and ABP; ECG Lead I; ECG Lead II; ECG Lead V; or ABP

Table 6 *P* values of t tests between errors of complexity means (manual vs. detected)

(p value)		Lead I	Lead II	Lead V	ABP
	Error	0.04 ± 0.47	0.05 ± 0.33	0.2 ± 0.34	0.57 ± 0.68
Fusion	−0.01 ± 0.35	$p = 0.01$	$p = 0.02$	$p = 0.4$	$p < 0.0001$

sequences and those derived from automatically detected peaks) showed that the "Fusion" values were the least statistically different from the gold standard (see Table 4). Furthermore, using 0.1 as the difference considered practically zero, equivalence tests showed that only the "Fusion" values were practically equivalent to the gold standard (see Table 5). Lastly, the fusion of waveform sources produced better error density distributions than those derived from individual waveforms (see Fig. 3) as well as narrower confidence intervals. Statistical significances were determined for all error comparisons ($p < 0.05$), except between "Fusion" and Lead V ($p = 0.35$), with the most significance between the "fusion" of waveform sources and ABP ($p < 0.001$).

4 Discussion

Although this study involved a fairly standard dataset [30] based upon availability of three ECG lead as well as arterial blood pressure waveforms, the dataset presented enough noisy waveforms to challenge our data fusion framework. Therefore, due to the quality of underlying data, the performance of R-wave detection reported in Table 3 was quite low. Nevertheless, better R-wave detection performance results in better signal-derived metrics (see Tables 3, 4, and 5). (It is important to note here that many previously published results of beat detection algorithms against different databases [e.g., MIT–BIH Arrhythmia Database] involved the detection of beats or

QRS complexes, rather than detection of R-waves. Hence, their results may not reflect stringent requirements on performance and may be better suited for beat applications, as opposed to the real-time calculation of HRC.)

An initial analysis was conducted to observe whether a fusion of three inferior waveforms produced results commensurate with those obtained from one reliable waveform (e.g., ECG Lead II); afterwards, data analysis involved all four waveforms (Leads I, II, V, and ABP). Importantly, the fusion of all available waveforms produced results better than those obtained from one reliable waveform (e.g., ECG Lead II). Visual comparisons of the error density distributions (histograms) superimposed on the left-hand sides of each Bland–Altman plot were made in order to determine whether the data fusion framework yielded the best results. The fusion of waveform sources produced slightly better error density distributions than those belonging to individual waveform sources (see Fig. 3; Table 6). Therefore, our method for calculating HRC has much potential application in the clinical environment.

Statistical significances were found for all error comparisons ($p < 0.05$), except between "Fusion" and Lead V ($p = 0.35$), with the most significance between ABP and manual verification ($p < 0.001$), suggesting that the ECG and ABP waveforms of patient records in the MGH/MF Waveform Database are disparate in quality. The fact that errors for Lead V were not significantly different from those for the fusion of waveform channels also demonstrated that the dominant lead (in this case, Lead II) may not always be reliable for signal-derived metrics. As expected, because of the ABP waveform's poor quality, our data fusion framework rejected use of this signal during fusion; this proved valuable for testing the framework. Had all waveforms been degraded in quality, a data fusion framework could prove optimal for calculating HRC in the ICU environment.

The Bland–Altman plots also show different patient groups through respective symbols and colors. Green squares denote patients with NSR, while magenta diamonds denote patients with ST. Blue circles denote patients with AF. From these plots, patients with NSR have varying complexity values and an overall mean entropy greater than that for patients with ST, implying that the latter group often involves patients with concomitant illnesses or pathologies. These observations agree with findings in [19–22]. Equivalence tests demonstrated the data fusion framework's overall reliability. While errors for Lead V were not significantly different from those for the fusion of waveform channels, this was not as apparent when considering equivalence between mean values. Because mean HRC values varied by rhythm, i.e., AF (1.0 ± 0.9); NSR (0.9 ± 0.5); and ST (0.7 ± 0.5), from Tables 4 and 5, this study showed that overall improved HRC calculations can better discriminate patient groups.

From our experiments involving standard test data as well as laboratory data at the Institute of Surgical Research, we observed that the four individual PD algorithms that compose the AESOP and BEBOP algorithms each demonstrated particular strengths and weaknesses, thus reaffirming the use of a fusion algorithm for detecting peaks within an ECG or non-ECG signal, respectively, especially in a real-time, ambulatory environment. For specific records, each component algorithm functioned better than the other algorithms according to our implementations.

In light of recent work by Moorman et al. [31, 32] and Seely et al. [33, 34] investigating the clinical use of HRC and heart-related metrics for detecting sepsis and multiorgan failure, improvement of HRC calculations may help detect significant changes from baseline values earlier and more accurately. Moreover, improved HRC calculations could help improve trends over seconds, minutes, or even days and identify crossed thresholds that would have otherwise been missed due to poor RWD performance. Consequently, improved HRC accuracy could enhance clinical decision making as well as decision support systems in the areas mentioned above. Other applications which require monitoring over time, such as mentioned in [35], could likewise benefit from improvements in HRC values.

Our work may also be extended to include waveforms derived from pulse oximetry, a more common and non-invasive method of measuring the pulsatile flow in the cardiovascular system, Furthermore, the data fusion framework described above may provide a general and practical solution for extracting any heart-related characteristics in the critical care environment.

Two limitations of this study were that (1) the MGH/MF Waveform Database involved a dominant ECG lead for many records and (2) annotated files containing manually verified R waves depended on one reference signal. A future study may be for us to run analyses on waveforms with known amounts of degraded data, thereby demonstrating not only reliability of HRC calculations but also improved robustness. We also hope that future work using a different dataset may better address the issue of clinical outcomes, as related to improved R-wave detection performance and heart-rate complexity calculations.

5 Conclusion

Two fusion algorithms (AESOP and BEBOP) were developed for detecting the peaks of ECG and non-ECG waveforms, respectively. They were then incorporated into a framework and real-time system for calculating HRC using multiple waveforms and signal quality indices. The data fusion framework was shown to provide in real-time a

reliable continuously streamed HRC value, derived from multiple waveforms in the presence of noise and artifacts. This approach will be validated and tested for assessment of HRC in trauma patients.

Acknowledgments This work was supported by the U.S. Army Institute of Surgical Research Comprehensive Intensive Care Research Task Area and the U.S. Army Combat Casualty Care Research Program. We would like to thank Mr. George B. Moody and Mr. Ken Pierce of PhysioNet for converting a rare printed copy of an MGH/MF Waveform Database patient guide into computer-readable form in order to make its contents available and searchable.

Conflict of interest The authors declare that they have no conflict of interest.

Disclaimer The opinions or assertions contained herein are the private views of the authors and are not to be construed as official or as reflecting the views of the Department of the Army or Department of Defense.

References

1. Batchinsky AI, Cooke WH, Kuusela T, Cancio LC. Loss of complexity characterizes the heart-rate response to experimental hemorrhagic shock in swine. Crit Care Med. 2007;35:519–25.
2. Batchinsky AI, Cancio LC, Salinas J, Kuusela T, Cooke WH, Wang JJ, Boehme M, Convertino VA, Holcomb JB. Prehospital loss of R-to-R interval complexity is associated with mortality in trauma patients. J Trauma. 2007;63:512–8.
3. Cancio LC, Batchinsky AI, Salinas J, Kuusela T, Convertino VA, Wade CE, Holcomb JB. Heart-rate complexity for prediction of prehospital lifesaving interventions in trauma patients. J Trauma. 2008;65:813–9.
4. Batchinsky AI, Salinas J, Kuusela T, et al. Rapid prediction of trauma-patient survival by analysis of heart-rate complexity: impact of reducing dataset size. Shock. 2009;32:565–71.
5. Batchinsky AI, Skinner J, Necsoiu C, et al. New measures of heart-rate complexity: effect of chest trauma and hemorrhage. J Trauma. 2010;68:1178–85.
6. Li Q, Mark RG, Clifford GD. Artificial arterial blood pressure artifact models and an evaluation of a robust blood pressure and heart rate estimator. Biomed Eng Online. 2009;8:1–15.
7. Li Q, Mark RG, Clifford GD. Robust heart rate estimation from multiple asynchronous noisy sources using signal quality indices and a Kalman filter. Physiol Meas. 2008;29:15–32.
8. Ebrahim MH, Feldman JM, Bar-Kana I. A robust sensor fusion method for heart rate estimation. J Clin Monit. 1997;13:385–93.
9. Feldman JM, Ebrahim MH, Bar-Kana I. Robust sensor fusion improves heart rate estimation: clinical evaluation. J Clin Monit. 1997;13:379–84.
10. Tarassenko L, Mason L, Townsend N. Multi-sensor fusion for robust computation of breathing rate. Electron Lett. 2002;38:1314–6.
11. Pan J, Tompkins WJ. A real-time QRS detection algorithm. IEEE Trans Biomed Eng. 1985;32:230–6.
12. Hamilton PS, Tompkins WJ. Quantitative investigation of QRS detection rules using the MIT/BIH arrhythmia database. IEEE Trans Biomed Eng. 1986;33:1157–65.
13. Christov II. Real time electrocardiogram QRS detection using combined adaptive threshold. BioMed Eng Online. 2204;3:1–9.
14. Afonso VX, Tompkins WJ, Nguyen TQ, et al. ECG beat detection using filter banks. IEEE Trans Biomed Eng. 1999;46:192–202.
15. Zong W, Moody GB, Jiang D. A robust open-source algorithm to detect onset and duration of QRS complexes. Comput Cardiol. 2003;30:737–40.
16. Friesen GM, Jannett TC, Jadallah MA, et al. A comparison of the noise sensitivity of nine QRS detection algorithms. IEEE Trans Biomed Eng. 1990;37:85–98.
17. Richman JS, Moorman JR. Physiological time series analysis using approximate entropy and sample entropy. Am J Physiol Heart Circ Physiol. 2000;278:H2039–49.
18. Lake DE, Richman JS, Griffin MP, Moorman JR. Sample entropy analysis of neonatal heart rate variability. Am J Physiol Regul Integr Comp Physiol. 2002;283:R789–97.
19. Griffin MP, Moorman JR. Toward the early diagnosis of neonatal sepsis and sepsis-like illness using novel heart rate analysis. Pediatrics. 2001;107:97–104.
20. Griffen PM, O'Shea TM, Bissonette EA, Harrell FE, Lake DE, Moorman JR. Abnormal heart rate characteristics preceding neonatal sepsis and sepsis-like illness. Pediatr Res. 2003;53:920–6.
21. Costa M, Goldberger AL, Peng C-K. Multiscale entropy analysis of biological signals. Phys Rev E. 2005;71:021906-1–021906-18.
22. Costa M, Goldberger AL, Peng C-K. Multiscale entropy analysis of physiologic time series. Phys Rev Lett. 2002;89:068102-1–068102-4.
23. Acharya UR, Joseph KP, Kannathal N, Lim CM, Suri JS. Heart rate variability: a review. Med Bio Eng Comput. 2006;44:1031–51.
24. Malik M, Camm AJ. Heart rate variability: from facts to fancies. J Am Coll Cardiol. 1993;22:566–8.
25. Task Force of the European Society of Cardiology and the North American Society of Pacing and Electrophysiology. Heart rate variability. Standards of measurement, physiological interpretation and clinical use. Circulation. 1996;93:1043–65.
26. Welch JP, Ford PJ, Teplick RS, et al. The Massachusetts general hospital-Marquette foundation hemodynamic and electrocardiographic database comprehensive collection of critical care waveforms. J Clin Monit. 1991;7:96–7.
27. Goldberger AL, Amaral LAN, Glass L, et al. PhysioBank, PhysioToolkit, and PhysioNet: components of a new research resource for complex physiologic signals. Circulation. 2000;101:215–20.
28. Moody GB, Mark RG, Goldberger AL. PhysioNet: a web based resource for the study of physiologic signals. IEEE Eng in Med and Biol. 2001;20:70–5.
29. Liu NT, Batchinsky AI, Cancio LC, Baker WL, Salinas J. Development and validation of a novel fusion algorithm for continuous, accurate, and automated R-wave detection and calculation of signal-derived metrics. J Crit Care. 2013. doi:10.1016/j.jcrc.2013.02.015.
30. Köhler B-U, Hennig C, Orglmeister R. The principles of software QRS detection. IEEE Eng Med Biol Mag. 2002;21:42–57.
31. Moorman JR, Carlo WA, Kattwinkel J, et al. Mortality reduction by heart rate characteristic monitoring in very low birth weight neonates: a randomized trial. J Pediatr. 2011;6:900–6.
32. Moorman JR, Delos JB, Flower AA, et al. Cardiovascular oscillations at the bedside: early diagnosis of neonatal sepsis using heart rate characteristics monitoring. Physiol Meas. 2011;32:1821–31.
33. Green IG, Bradley B, Bravi A et al. Continuous multiorgan variability analysis to track severity of organ failure in critically ill patients. J Crit Care. 2013. doi:10.1016/j.jcrc.2013.04.001.
34. Bravi A, Green G, Longtin A, et al. Monitoring and identification of sepsis development through a composite measure of heart rate variability. PLoS One. 2012;7(9):e45666.
35. Bravi A, Longtin A, Seely AJE. Review and classification of variability analysis techniques for clinical applications. BioMed Eng Online. 2011;10:1–27.